

Parallelization of Minimal Problem Solver Generator

Bc. Vojtěcha Cvrčka

Diplomová práce Vojtěcha Cvrčka se zabývá paralelizací výpočtů, které jsou součástí robustního odhadu parametrů esenciální matice a geometrické verifikace předběžných korespondencí. Předpokládá se, že tyto výpočty jsou implementovány ve schématu daném metodou RANSAC. Zde jsou potřebné dva moduly: jednak výpočet parametrů esenciální matice z minimálního vzorku korespondencí, který vyžaduje tzv. *minimal problem solver* a dále nalezení podpory pro tuto instanci v souboru korespondencí (*support*). Minimal problem solver hledá řešení soustavy polynomiálních rovnic. Důležitým krokem je nalezení vlastních čísel a vlastních vektorů nederogatorní čtvercové matice, jejíž rozměr je například pro pětibodový algoritmus 10×10 . Zjištění podpory nalezených parametrů vyžaduje výpočet Sampsonovy chyby pro všechny předběžné korespondence, což je obvykle výpočetně drahá operace, neboť počet předběžných korespondencí roste kvadraticky s počtem detekovaných bodů zájmu a je pro standardní problém řádově 10 000. To je technický problém pro úlohy *structure-from-motion*, které mají běžet v reálném čase. Cíl zrychlit tyto výpočty je tedy legitimní a úspěšná řešení jsou pro obor počítačového vidění zajímavá.

Předložená práce se zabývá jednotlivými částmi výpočtu a navrhuje implementaci na grafické kartě v prostředí CUDA. Dosažené výsledky ukazují, že paralelizace výpočtu podpory parametrů (*support*) dosahuje významného zrychlení. Podle mého názoru ale implementace tohoto výpočtu není náročná. Zajímavější je pokus o paralelizaci výpočtu vlastních čísel a vektorů. Diplomant prostudoval známé metody navržené pro neparalelní výpočet a nakonec použil algoritmus popsany v knize [12], s modifikacemi ze zdroje [3]. Využil přitom speciálních vlastností nederogatorních matic. Tento algoritmus implementoval v CUDA. Z popisu v práci uvedeného však není příliš zřejmé jak náročná tato implementace byla. Experimentální výsledky nenaznačují, že by takováto implementace mohla být lepší než implementace používající vhodnou metodu paralelizace implementovanou na CPU. To je v souladu s tím, co je známo v komunitě oboru počítačového vidění.

Potud se zdá, že zadání diplomového úkolu bylo splněno. Diplomant nastudoval nesnadné partie numerických metod, vybral metodu vhodnou k paralelní implementaci, po úpravách ji naimplementoval a experimentálně ověřil její výkonnost.

Provedení předložené práce ale nespĺňuje standardy kladené na odborný výklad a dokumentaci experimentální práce výše uvedeného typu. Čtenář vůbec není uveden do problému a musí práci číst tak, že průběžně studuje odkazované zdroje, zejména [12], aby vůbec mohl postoupit dále. To při studiu odbodného textu sice musí dělat vždy, zde je ale tato potřeba nad únosnou míru. Řazení kapitol je chybné, text je nejen velmi lakonický, ale též značně nepřehledný a chaotický, chybí vysvětlení značného počtu odborných termínů, terminologie kolísá, ve vzorcích a popisech algoritmů jsou závažné chyby a převzaté výsledky týkající se algoritmů jsou zkráceny, některé ilustrace nemají vůbec popisku, chybí popis experimentů a jejich výsledky nejsou komentovány a čtenář musí často hádat co má autor na mysli. V práci je mnoho překlepů a gramatických chyb. Práce činí dojem, že byla napsána za několik hodin těsně před termínem.

K práci mám tyto konkrétní výhrady:

1. Text práce nesleduje standardní strukturu dokumentu (IMRAD):
 - (a) práce začíná popisem CUDA (kap. 1), což je technikalita, místo toho, aby se nejprve popsal problém, který bude řešen (to si čtenář domýšlí z poznámek rozptýlených v textu).

- (b) Dále například popis *state of the art* je až v kapitole 5, místo toho, aby byl součástí úvodu.
2. Popis a diskuse jsou místy nedostatečné:
- (a) Čtenáři značně ztěžuje čtení například to, že musí uhádnout co přesně je *automatic solver* a co je *solver generator*. To bez detailní předchozí znalosti problematiky není možné.
 - (b) Mimochodem, není jasné, proč je v názvu práce slovo *Generator*. Práce o generátoru není. Je to jen chyba z nepozornosti?
 - (c) Co dělá *aggressive early deflation* zmíněná jako důležitá metoda v kap. 4.5? Používá se tato metoda v popsané implementaci?
 - (d) Nerozumím Alg. 2 na str. 21. Co je vstupem? Co je $H_{n,n}$ a čím se liší od $h_{n,n}$ ve (4.12)? Co je H a proč je algoritmem přepsáno?
 - (e) Co je to *pair of points* v kap. 6.3.1? Korespondence?
 - (f) Popis implementace triangulace (kap. 6.3.1) není dostatečný, chybí vzorce, ze kterých by bylo zřejmé, jaký je postup.
 - (g) V popisu implementace Sampsonovy chyby na str. 29 se píše, že implementace se neliší. Od čeho?
 - (h) Která metoda akumulace počtu *inliers* ze dvou zmíněných v kap. 6.3.2 byla použita?
 - (i) V kap. 6.5 a obr. 6.3 není zřejmé proč je v paměti m matic. Co je m ? Zřejmě je to jiné m než je uvedeno v popisu algoritmu na str. 23.
 - (j) Bývá zvykem, že pokud ve vzorci uvedeme málo používanou funkci, jako například nejmenší společný násobek $lcm(\cdot, \cdot)$, tak ji vysvětlíme, viz (6.1).
 - (k) Co vidíme v obr. 7.9 – 7.13? Popis těchto i ostatních experimentálních výsledků zcela chybí.
3. V práci je značné množství chyb.
- (a) Algoritmus 1 na str. 18 nedává žádný smysl. Při pohledu do zdroje [12] vidíme, že druhý řádek cyklu má počítat T_k , nikoliv T_{k-1} .
 - (b) Výraz (4.10) se použije pro výpočet y_{ij} , nikoliv pro výpočet y_j , jak uvedeno.
 - (c) Tvrzení v kap. 4.10 je nesrozumitelné. Při pohledu do [12, p. 387] se ukazuje, že je zkrácen význam původního tvrzení.
 - (d) Alg. 4 je chybně vysázený. Pohled do [12 p. 394] odhaluje původní verzi. Co je vstupem a co je výsledkem tohoto algoritmu?
 - (e) Co vidíme v obr. 7.8? Jsou to výsledky pro obě datové sady? Pro jednu? Kterou?
4. Terminologie je neustálená nebo nepřesná.
- (a) Proč se termín *problem solver* nevyskytuje v popisu RANSACu na str. 16? Pouze se zmiňuje, že je nutno *calculate model parameters*.
 - (b) V angličtině používáme výraz *asymmetric matrix*, nikoliv *unsymmetric matrix*.
 - (c) V kontextu RANSACu bychom neměli používat výraz *model* pro *proposal* a *model testing* pro *determining support*. Ale pravda je, že tato terminologie je i v literatuře rozkolísaná vlivem nepoučených autorů. Na druhou stranu s terminologií používanou v diplomové práci jsem se nesetkal. O to více je nutné v technickém textu přesně vysvětlit s jakými koncepty pracujeme a nenechat čtenáře hádat. Nepřispívá například ani to, že to, čemu se v kap. 3.3 a 6 říká *model test* se v kap. 7 nazývá *verification*.

5. Práce se prohřešuje proti zavedenému technickému stylu:
 - (a) Číslované odkazy, například Algorithm 1, Figure 6.1, Theorem 2, Section 4.8 se v angličtině píše s velkým písmenem a bez členu.
 - (b) Vkládání obrázků a výpisů programů *inline* je nestandardní a značně tříští a zne-
přehledňuje text. Proč nejsou tyto prvky vysázeny jako *floats*?
 - (c) V práci je značné množství krátkých odstavců, které mají samostatnou číslovanou
podkapitulu, např. 1.2, 1.3, 2.1.1, 2.1.2, 4.5, atd, atd.
6. Práce se zdroji a odkazování zdrojů by mohly být lepší:
 - (a) Například pokud přebíráme Alg. 1 ze zdroje [12] a tento zdroj je kniha, je zvy-
kem doplnit referenci o číslo kapitoly nebo stránky takto: [12, p. 376]. L^AT_EX pro to
poskytuje nástroje.
 - (b) Není jasné, jak dalece autoritativním zdrojem je přednáška [3]. Není k dispozici
původní zdroj?
 - (c) Reference [33] je chybně zformátována. V ostatních referencích se často nerespektuje,
že zkratky musí být velkými písmeny a vlastní jména začínají velkým písmenem.

Dále mám několik dotazů:

1. Jsou 2×2 diagonální bloky ve Schurově dekompozici horní trojúhelníkové, aby bylo možné
použít *decoupling*, jak popsáno v kap. 6.6.2?
2. *Inverse iteration* popsaná v Alg. 4 je iterativní metoda. Jak je iterace implementována?
Kolik iterací se v popsané implementaci provede?
3. Jak byly vybrány datové sady pro experimenty v kap. 7.4 a jaké měly tyto sady vlastnosti?
4. Proč byly v kap 7.4 výsledky celého RANSACu hodnoceny jen na dvou datových sadách?
Jsou odchylky mezi sadami malé nebo velké?

Kvalita diplomové práce se posuzuje také podle kvality předloženého textu. Ten je stěží
uspokojivý. Při hodnocení přesnosti popisu algoritmů beru v úvahu to, že některé nedostatky
v popisu algoritmu jsou způsobeny nekritickým přenosem z knihy [12], kde je notace v zájmu
stručnosti úmyslně zkrácena, ovšem nad obvyklou míru. Vzhledem k tomu, že zadání bylo
splněno, doporučuji práci k prezentaci a navrhuji hodnocení stupněm **C – dobře**.

Doc. Dr. Techn. Ing. Radim Šára
katedra kybernetiky FEL ČVUT
6. 6. 2016